# Single Server Systems — I. Relations Between Some Averages

By S. O. RICE

(Manuscript received April 13, 1961)

*This is the first of two papers dealing with single server systems. Two subjects are discussed in the present paper, (i) relations between such items as the probability of loss, probability of no delay, and the average number of customers served in a busy period, and (ii) the statistical behavior of a single server system in which no waiting for service is allowed.*

## I. INTRODUCTION

A typical but rather homely example of the systems considered here is a barber shop in which there is only one barber, the "single server." Let $N - 1$ be the number of chairs provided for customers waiting for service so that the "capacity" of the system is $N$. When the shop is full, one customer is being served and $N - 1$ are waiting. A prospective customer (a "demand" for service) arriving when the shop is full is turned away and is said to be "lost." If the shop is not full, he waits and is eventually served.

The demands arrive at an average rate of $a$ per unit time. The server would serve $b$ customers per unit time (on the average) if he were to work steadily. It follows that the average interval between arrivals is $1/a$ and the average service time is $1/b$. It is assumed that the rates $a$ and $b$ do not change with time.

The first part of the paper is concerned with several quantities of interest, including the fraction $L$ of demands lost and the average length of the busy periods, i.e., the periods during which the server is continuously busy. The values of these quantities are expressed in terms of $a$, $b$, and two other quantities $p_0$ and $\tau$. Here $p_0$ is the probability that the server is idle (at an instant selected at random) and $\tau$ is the average duration of an idle period. Both $p_0$ and $\tau$ depend upon $N$ and upon the probability laws governing the arrivals and service times. However, only the simplest cases of this dependence are mentioned in the first part of the paper. The results are summarized in Table I.

The second part of the paper is concerned with the single server "loss system" in which no waiting is allowed ($N = 1$). The input is now assumed to be "recurrent," i.e., the distances between arrivals are independent and have the general distribution function $A(t)$. The service times have the distribution function $B(t)$ and are independent of each other and of the arrivals. The functions $A(t)$ and $B(t)$ are such that the interarrival distances and service lengths have the respective average values

$$a^{-1} = \int_0^\infty [1 - A(t)]\, dt$$
$$b^{-1} = \int_0^\infty [1 - B(t)]\, dt. \tag{1}$$

Further conditions are imposed on $A(t)$ and $B(t)$ in the course of the derivations. The principal items of interest are (*i*) the loss $L$, (*ii*) the probabilities $p_0$ and $p_1$ that the server is idle or busy, respectively, at a time selected at random, and (*iii*) the distribution of the lengths of the idle periods. The expression for the loss is equivalent to one obtained in a different manner by F. Pollaczek.[1] A discussion of how the loss increases with the variability of arrival has been given by P. M. Morse.[2]

Results for a Type I counter, i.e., one in which the registration of an arrival is followed by a "dead time," may be applied to the single server loss system by identifying the dead time (supposed variable) with the intervals the server is busy. For example, results obtained here are closely related to some for counters given by R. Pyke[3] and earlier workers to whom he makes reference.

The infinite capacity system ($N = \infty$) is discussed in a companion paper.[4] In this case there is no loss, and attention is focused on the distributions of the waiting times and busy period lengths. Some of the results of Section II of the present paper find application there.

I am indebted to John Riordan for many helpful discussions on the subject matter of these two papers and for numerous improvements in presentation.

## II. AVERAGES OBTAINED FROM FIRST PRINCIPLES, GENERAL INPUT AND LIMITED CAPACITY

Several important averages associated with a single server system may be readily obtained by considering its behavior over a long period of time $T$, statistical equilibrium being assumed, and then letting $T \to \infty$. The input and service are assumed to be general with respective arrival

and service rates $a$ and $b$. The capacity of the system is $N$, so that demands arriving when $N - 1$ are waiting are lost.

The probability $p_0$ that the server is idle (at an instant selected at random) and the average duration $\tau$ of an idle period are supposed given. For Poisson input, i.e., one in which the probability that a demand will arrive in the interval $t$, $t + dt$ is $adt + 0(dt^2)$ (irrespective of the arrival times of the other demands), the value of $\tau$ is $1/a$.

Let $\nu(T)$ be the number of arrivals in $T$ and let $\epsilon$ be a given, arbitrarily small, positive number. The input and service are assumed to be such that the probability of $1 - \epsilon < \nu(T)/aT < 1 + \epsilon$ may be made as close to unity as desired by choosing $T$ large enough. For recurrent input this restriction is satisfied, by virtue of the law of large numbers, when the first integral in (1) gives a finite value for $a^{-1}$. We shall refer to the above inequality by saying that the number of arrivals in the long interval $T$ is "equal" to $aT$. Similar statements made below regarding total idle and busy time, number served, etc., in the interval $T$ are to be interpreted in a similar way.

The total idle time of the server is $Tp_0$, the total busy time is $T - Tp_0$, and the total number served is $b(T - Tp_0)$. The number lost is the number of arrivals less the number served, and the fraction of arrivals lost is

$$L = [aT - b(T - Tp_0)]/aT = 1 - (1 - p_0)\rho^{-1} \tag{2}$$

where $\rho = a/b$. $L$ is the "probability of loss." Loss occurs only when the waiting room is filled, i.e., when the system is in state $N$ (probability $p_N$ and average duration $\tau_N$). The number of times state $N$ occurs is $p_N T/\tau_N$, and the average number of demands lost during such a period is $\tau_N[a - b(1 - p_0)]/p_N$. For exponential service the service time lengths have the distribution function $B(t) = 1 - e^{-bt}$, and the value of $\tau_N$ is $1/b$.

The number served in $T$ without delay is equal, to within one, to the number of idle periods. The number of idle periods is $p_0 T/\tau$; hence $p_0/a\tau$ is the probability $W(0)$ that a demand will be served upon arrival. Thus,

$$W(0) = p_0/a\tau. \tag{3}$$

Because a demand is either served without delay, delayed, or lost, the probability of delay is

$$1 - L - (p_0/a\tau) = (1 - p_0)\rho^{-1} - (p_0/a\tau).$$

Since busy and idle periods alternate, the number of busy periods is, to within one, the number of idle periods, namely $p_0 T/\tau$. Dividing this

into the total busy time $T - Tp_0$ gives $\tau(1 - p_0)/p_0$ for the average length of a busy period. Similarly, the total number served gives $b\tau(1 - p_0)/p_0$ for the average number served in a busy period.

When $N$ is infinite, statistical equilibrium requires $\rho < 1$, and all demands are eventually served. Equating the number of arrivals $aT$ to the number served $b(T - Tp_0)$ gives $p_0 = 1 - \rho$.

These results and the forms they assume for $N$ infinite and $\tau = 1/a$ (Poisson input) are summarized in Table I. The results for Poisson input are well known.

When the capacity of the waiting room is infinite, the average number $\bar{n}$ of demands served in a busy period is

$$\bar{n} = 1/W(0). \tag{4}$$

In this case $L$ is zero, demands are either delayed or not delayed, and $W(0)$ becomes the probability of no delay. This follows from (3), Table I, and the fact that both sides of (4) are equal to $a\tau/(1 - \rho)$. For limited capacity

$$\bar{n} = (1 - L)/W(0) \tag{5}$$

which follows from Table I. Here $W(0)$ is the chance that a demand chosen at random will be served upon arrival.

## III. SINGLE SERVER LOSS SYSTEM

In a single server loss system any demand arriving when the server is busy is lost. The system capacity $N$ is 1, there is no waiting line, and every busy period consists of a single service interval. As mentioned in the introduction, the input is assumed to be recurrent, the service is

TABLE I — STATIONARY AVERAGES FOR A GENERAL SYSTEM

| Average | Limited Capacity General Input | Infinite Capacity, $p_0 = 1 - \rho$ | |
|---|---|---|---|
| | | General Input | Poisson Input, $\tau = a^{-1}$ |
| Rel. number lost | $1 - (1 - p_0)\rho^{-1}$ | 0 | 0 |
| Rel. number not delayed | $p_0(a\tau)^{-1}$ | $(1 - \rho)(a\tau)^{-1}$ | $1 - \rho$ |
| Rel. number delayed | $(1 - p_0)\rho^{-1} - p_0(a\tau)^{-1}$ | $1 - (1 - \rho)(a\tau)^{-1}$ | $\rho$ |
| Length of busy period | $\tau(1 - p_0)p_0^{-1}$ | $\tau\rho(1 - \rho)^{-1}$ | $(b - a)^{-1}$ |
| Number served in busy period | $b\tau(1 - p_0)p_0^{-1}$ | $a\tau(1 - \rho)^{-1}$ | $(1 - \rho)^{-1}$ |

Notation: $a$ = arrival rate, $b$ = service rate, $\rho = a/b$, $\tau$ = average length of idle period, $p_0$ = fraction of time server is idle.

general, and the principal items of interest are (*i*) the loss $L$, (*ii*) the probabilities $p_0$ and $p_1$ that the server is idle or busy, respectively, at a time selected at random, and (*iii*) the probability $q(u)\,du$ that the length of an idle period will lie between $u$ and $u + du$.

3.1 *Values of $L$, $p_0$ and $p_1$*

For general input and service, the results of Section II for $N = 1$ and the relation $1 - p_0 = p_1$ show that

$$L = 1 - p_1\rho^{-1}, \qquad p_0 = b\tau p_1, \qquad \rho = a/b \tag{6}$$

where the second relation is obtained by equating the average busy period length $\tau(1 - p_0)/p_0$ to the average service length $1/b$, $\tau$ being the average idle period length.

For Poisson input, $\tau = 1/a$ and (6) gives

$$p_1 = \rho p_0, \qquad p_0 = (1 + \rho)^{-1}, \qquad L = p_1 = \rho/(1 + \rho). \tag{7}$$

In this case, the loss is independent of the service distribution, a property also possessed by the many-server loss system for Poisson input.

For recurrent input and general service, the ratio $L/(1 - L)$ of the number of demands lost to the number served is equal to the expected number of demands arriving while the server is busy serving one demand, i.e., during one service interval. Thus

$$\frac{L}{1 - L} = \int_0^\infty \overline{n(t)}\, dB(t) \tag{8}$$

where $\overline{n(t)}$ is the expected number of arrivals during a service of length $t$ (not counting the one starting the service) and $B(t)$ is the service time distribution function. As $t$ becomes large, $\overline{n(t)}$ is $0(at)$ and the integral converges because the average service time is finite.

It will be shown that

$$\overline{n(t)} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{\alpha(s)e^{st}\,ds}{[1 - \alpha(s)]s} \qquad (c > 0) \tag{9}$$

where $\alpha(s)$ is the Laplace-Stieltjes transform of the distribution function $A(t)$ for the separation between arrivals:

$$\alpha(s) = \int_0^\infty e^{-st}\, dA(t). \tag{10}$$

It is assumed that $\alpha(s)$ and $t$ are such that the integral in (9) converges. Equation (9) is but one of a number of similar results; see, for example, D. R. Cox and W. L. Smith.[5]

To obtain (9), note that the service starts with an arrival, and the probability that $n$ or more additional arrivals will occur in the ensuing interval of length $t$ is the probability that

$$S_n \equiv X_1 + X_2 + \cdots + X_n \leqq t. \tag{11}$$

Here $X_i$ is the separation between arrivals $i - 1$ and $i$. Since the $X_i$'s are independent and have the distribution function $A(t)$, the rules for determining the distribution of the sum of $n$ random variables may be applied to find the chance that $S_n \leqq t$. In particular, the Laplace transform of the probability density for $S_n$ is $[\alpha(s)]^n$, and the Laplace transform of Prob $[S_n \leqq t]$ is $[\alpha(s)]^n/s$.

The probability of exactly $n$ arrivals in an interval of length $t$ which starts just after an arrival is

$$\begin{aligned} P_n(t) &= \text{Prob}\,[S_n \leqq t) - \text{Prob}\,[S_{n+1} \leqq t] \\ &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} [1 - \alpha(s)][\alpha(s)]^n s^{-1} e^{st}\, ds. \end{aligned} \tag{12}$$

Multiplying $P_n(t)$ by $n$, noting that $|\alpha(s)| < 1$ on the path of integration, and summing from $n = 0$ to $n = \infty$ then gives (9).

When $\overline{n(t)}$ is known as a function of $t$, either from (9) or otherwise, and is used in (8), the result is an equation which may be solved for the loss $L$. When $L$ is known, (6) gives

$$p_1 = \rho(1 - L), \qquad p_0 = 1 - p_1\,. \tag{13}$$

A few examples follow.

*i.* Poisson input. Here $1 - A(t) = e^{-at}$ and $\alpha(s) = a/(a + s)$. Then

$$\overline{n(t)} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} as^{-2} e^{st}\, ds = at$$

$$\frac{L}{1 - L} = \int_0^\infty at\, dB(t) = a/b = \rho$$

and the results given in (7) are again obtained.

*ii.* Arrivals spaced $1/a$ apart. By inspection

$$\overline{n(t)} = n, \qquad na^{-1} < t < (n + 1)a^{-1}$$

and (8) becomes

$$\frac{L}{1 - L} = \sum_0^\infty n \int_{n/a}^{(n+1)/a} dB(t) = \sum_1^\infty n\left[B\left(\frac{n+1}{a}\right) - B\left(\frac{n}{a}\right)\right]. \tag{14}$$

*iii.* When $\alpha(s)$ is $0(1/s)$ as $s \to c \pm i\infty$, as it is when the probability density $A'(t) = dA(t)/dt$ exists and is of bounded variation, and when

$B'(t)$ exists and is $0(e^{-\epsilon t})$, $\epsilon > 0$, as $t \to \infty$, substitution of (9) in (8) gives

$$\begin{aligned} \frac{L}{1-L} &= \int_0^\infty \frac{dB(t)}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\alpha(s)e^{st}\,ds}{[1-\alpha(s)]s} \\ &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\alpha(s)\beta(-s)\,ds}{[1-\alpha(s)]s} \qquad (0 < c < \epsilon) \end{aligned} \tag{15}$$

where

$$\beta(s) = \int_0^\infty e^{-st}\,dB(t). \tag{16}$$

In (15) the singularities of $\beta(-s)$ lie to the right of the path of integration and those of $\alpha(s)/s[1-\alpha(s)]$ to the left. The conditions imposed on $\alpha(s)$ and $B'(t)$ are sufficient to ensure the absolute convergence of the double integral and hence to justify the inversion of the order of integration. The result (15) is equivalent to one obtained by F. Pollaczek[1] by a different method.

*iv.* Exponential service. Here, $1 - B(t) = e^{-bt}$ and $\beta(-s) = b/(b-s)$. Substituting this value of $\beta(-s)$ in (15), closing the path of integration by an infinite semicircle on the right, and evaluating the residue at the pole $s = b$ gives

$$\frac{L}{1-L} = \frac{\alpha(b)}{1-\alpha(b)}, \qquad L = \alpha(b), \qquad p_1 = \rho[1-\alpha(b)].$$

This expression for $p_1$ is a special case of the stationary state probabilities (determined by both F. Pollaczek[6] and L. Takács[7]) for the many server system with recurrent input and exponential service.

*v.* Let $\alpha(s)$ satisfy the same condition as in example *iii* and in addition, suppose that (9) may be written as

$$\overline{n(t)} = R(t) + \frac{1}{2\pi i} \int_{-c-i\infty}^{-c+i\infty} \frac{\alpha(s)e^{st}\,ds}{[1-\alpha(s)]s} \qquad (c > 0)$$

where the only singularity of $e^{st}\alpha(s)/s[1-\alpha(s)]$ to the right of $\mathrm{Re}(s) = -c$ ($s$ finite) is a double pole at $s = 0$ with residue $R(t)$. Then (8) gives

$$\frac{L}{1-L} = \int_0^\infty R(t)\,dB(t) + \frac{1}{2\pi i} \int_{-c-i\infty}^{-c+i\infty} \frac{\alpha(s)\beta(-s)\,ds}{[1-\alpha(s)]s}. \tag{17}$$

It should be noticed that not all cases can be handled by (15) and (17). An example to the contrary is furnished by

$$\begin{aligned} A(t) &= 1 - (1+at)^{-2} \\ B(t) &= 1 - (1+bt)^{-2} \end{aligned} \tag{18}$$

where the averages $a^{-1}$ and $b^{-1}$ exist but the corresponding variances are infinite. In this case

$$\alpha(s) = 1 - x + x^2 e^x \int_x^\infty e^{-y} y^{-1}\, dy \qquad (x = s/a).$$

As $|s| \to \infty$ in $\mathrm{Re}(s) \geqq 0$, $\alpha(s)$ is $0(s^{-1})$. Near $s = 0$,

$$\alpha(s) = 1 - sa^{-1} - s^2 a^{-2} \ln s + 0(s^2).$$

Since $\beta(s)$ is similar to $\alpha(s)$, both $\beta(-s)$ and $\alpha(s)/s[1 - \alpha(s)]$ have branch points at $s = 0$. Equation (17) fails in this case because $s = 0$ is not a double pole. Equation (15) fails because it is impossible to draw a path of integration which separates the singularities of $\beta(-s)$ from those of $\alpha(s)/s[1 - \alpha(s)]$. When $A(t)$and $B(t)$ are given by (18) it seems necessary to work directly with (8) and (9), or else use some sort of limit.

### 3.2 *Lengths of Idle Periods*

Now turn to the distribution of $l_b$ and $l_i$, the lengths of a busy period and the following idle period, recurrent input and general service being assumed. The probability that $l_b \leqq t$ is simply $B(t)$, the service length distribution function. The distribution of the idle period length $l_i$ is more complicated. Its average length is, from (6),

$$r = p_0/bp_1 = a^{-1}(1 - L)^{-1} - b^{-1} \tag{19}$$

where $L$ is given by (8). The first step towards obtaining the probability $q(u)\, du$ that $u < l_i < u + du$ is to determine the conditional probability $q(u; t)\, du$ that $u < l_i < u + du$ given $l_b = t$.

Consideration of the arrival patterns which give no arrivals in $(t, t + u)$ followed by one in $(t + u, t + u + du)$ leads to

$$q(u; t)\, du = \sum_{n=0}^{\infty} \Pr\,[S_n < t; t + u < S_{n+1} < t + u + du] \tag{20}$$

where $S_0$ is 0 and $S_n$ for $n > 0$ is the sum (see (11)) of $n$ interarrival intervals $X_i$. An expression for the joint probability density of $S_n$ and $S_{n+1}$ may be obtained by inverting its double Laplace transform

$$\begin{aligned} &\text{ave exp } [-rS_n - sS_{n+1}] \\ &\qquad = \text{ave exp } [-(r + s)(X_1 + \cdots + X_n) - sX_{n+1}] \\ &\qquad = [\alpha(r + s)]^n \alpha(s). \end{aligned}$$

Integrating the density over the region $0 < S_n < t$, $u + t < S_{n+1} < u + t + du$ shows that the $n$th term, $n > 0$, in (20) is

$$\frac{du}{(2\pi i)^2} \int_{c-i\infty}^{c+i\infty} ds\, e^{s(u+t)} \alpha(s) \int_{c-i\infty}^{c+i\infty} [\alpha(r+s)]^n (e^{rt} - 1) r^{-1}\, dr \quad (21)$$

where $c > 0$.

Expression (21) holds only for $n > 0$. However, replacing the factor $(e^{rt} - 1)$ by $e^{rt}$ gives an expression which holds for $n \geqq 0$. Indeed, closing the path of $r$-integration on the right shows that the integral of $[\alpha(r + s)]^n r^{-1}$ is zero for $n > 0$. Closing it on the left shows that the integral of $e^{rt} r^{-1}$ is $2\pi i$ for $t > 0$ and leads to the correct value for $n = 0$.

Setting the modified form of expression (21) in the series (20) and performing the summation shows that $q(u; t)\, du$ is equal to an expression obtained by replacing $[\alpha(r + s)]^n (e^{rt} - 1)$ in (21) by $[1 - \alpha(r + s)]^{-1} e^{rt}$. The joint probability density of $l_b$ and $l_i$ is $q(u; t)B'(t)$ where $B'(t) = dB(t)/dt$. The probability density $q(u)$ is the integral of $q(u; t)B'(t)$ taken from $t = 0$ to $t = \infty$. Assuming $B'(t)$ to be $0(e^{-\epsilon t})$ as $t \to \infty$ and choosing the paths of integration $c \pm i\infty$ so that $0 < 2c < \epsilon$ makes the integral of $B'(t) \exp [t(s + r)]$ converge and have the value $\beta(-s - r)$. Changing the variable of integration from $r$ to $z = r + s$ and, for convenience in writing (22), taking $\alpha(s)$ to be such that the path of integration for $s$ may be shifted to $-\eta \pm i\infty$ (this implies that $A'(t)$ is $0(e^{-\eta t})$ as $t \to \infty$) gives finally

$$q(u) = \left(\frac{1}{2\pi i}\right)^2 \int_{-\eta-i\infty}^{-\eta+i\infty} ds\, e^{su} \alpha(s) \int_{\eta-i\infty}^{\eta+i\infty} \frac{\beta(-z)\, dz}{[1 - \alpha(z)](z - s)} \quad (22)$$

where $u \geqq 0$ and $\eta$ is an arbitrarily small positive number.

It may be shown that (22) reduces to a $e^{-au}$ for Poisson input, as it should, and to

$$q(u) = b[1 - \alpha(b)]^{-1} \int_0^\infty e^{-bv} A'(u + v)\, dv \quad (23)$$

for recurrent input and exponential service. Multiplying $q(u)$ by $\exp(-s'u)$, integrating $u$ from 0 to $\infty$, closing the path of integration for $s$ on the right, and dropping the prime from $s'$ shows that the Laplace transform of $q(u)$ is

$$\text{ave } e^{-su} = \text{ave } e^{-sl_i} = \frac{1}{2\pi i} \int_{\eta-i\infty}^{\eta+i\infty} \frac{\beta(-z)}{z - s} \left[\frac{\alpha(s) - \alpha(z)}{1 - \alpha(z)}\right] dz. \quad (24)$$

We also have

$$\text{ave } e^{-rl_b - sl_i} = \frac{1}{2\pi i} \int_{\eta - i\infty}^{\eta + i\infty} \frac{\beta(r - z)}{z - s} \left[ \frac{\alpha(s) - \alpha(z)}{1 - \alpha(z)} \right] dz \qquad (25)$$

which may be regarded as a double Laplace transform. In (24) and (25) the singularities of $\beta(-z)$ and $\beta(r - z)$ are supposed to lie to the right of the path of integration and the remaining singularities of the integrands to the left.

REFERENCES

1. Pollaczek, F., "Problems stochastiques," in *Memorial des Sciences Mathematiques*, Fasc. No. **136**, Gauthier Villars, Paris, 1957, p. 113, Eq. (9.40).
2. Morse, P. M., *Queues, Inventories and Maintenance*, John Wiley and Sons, New York, 1958, Chapter 5.
3. Pyke, R., Annals of Math. Stat., **29**, 1958, p. 737.
4. Rice, S. O., this issue, p. 279.
5. Cox, D. R., and Smith, W. L., Biometrika, **41**, 1954, p. 91.
6. Pollaczek, F., C. R. Acad. Sci., Paris, **236**, 1953, p. 1469.
7. Takács, L., Acta Math. Acad. Sci. Hungar., **7**, 1956, p. 419.